

**ANALISA PERBANDINGAN DATA MINING PADA  
KLASIFIKASI PENYAKIT JANTUNG MENGGUNAKAN  
ALGORITMA *EXTREME LEARNING MACHINE* (ELM) DAN  
*K-NEAREST NEIGHBOR* (K-NN)**



**Disusun sebagai salah satu syarat memperoleh Gelar Strata I  
pada Jurusan Informatika Fakultas Komunikasi dan Informatika**

**Oleh :**

**INTAN LARASATI**

**L200170091**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS KOMUNIKASI DAN INFORMATIKA  
UNIVERSITAS MUHAMMADIYAH SURAKARTA**

**2021**

**HALAMAN PERSETUJUAN**

**ANALISA PERBANDINGAN DATA MINING PADA  
KLASIFIKASI PENYAKIT JANTUNG MENGGUNAKAN  
ALGORITMA *EXTREME LEARNING MACHINE* (ELM) DAN *K-  
NEAREST NEIGHBOR* (K-NN)**

**PUBLIKASI ILMIAH**

oleh:

**INTAN LARASATI**

**L200170091**

Telah diperiksa dan disetujui untuk diuji oleh:

Dosen Pembimbing

A handwritten signature in blue ink, appearing to read 'Azizah', with a stylized flourish at the end.

**Azizah Fatmawati, S.T., M.Sc.**

**NIK.1198**

**HALAMAN PENGESAHAN**

**ANALISA PERBANDINGAN DATA MINING PADA  
KLASIFIKASI PENYAKIT JANTUNG MENGGUNAKAN  
ALGORITMA *EXTREME LEARNING MACHINE* (ELM) DAN *K-NEAREST NEIGHBOR* (K-NN)**

**OLEH  
INTAN LARASATI  
L200170091**

**Telah dipertahankan didepan Dewan Penguji  
Fakultas Komunikasi dan Informatika  
Universitas Muhammadiyah Surakarta  
Pada hari Sabtu, 26 Juni 2021**


**Dewan penguji:**

1. **Azizah Fatmawati, S.T., M.Sc.**  
(Ketua Dewan Penguji)
2. **Fatah Yasin Al Irsyadi, S.T., M.T.**  
(Anggota I Dewan Penguji)
3. **Dr.Eng. Yusuf Sulisty Nugroho, S.T., M.Eng.**  
(Anggota II Dewan Penguji)

  
(.....)  
  
(.....)  
  
(.....)

Dekan  
Fakultas Komunikasi dan Informatika



  
Nurgiyatna, S.T., M.Sc., Ph.D  
NIK.881

## **PERNYATAAN**

Dengan ini saya menyatakan bahwa dalam publikasi ilmiah ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali secara tertulis diacu dalam naskah dan disebutkan dalam daftar pustaka.

Apabila kelak terbukti ketidakbenaran dalam pernyataan saya diatas, maka akan saya pertanggungjawabkan sepenuhnya.

**Surakarta, 25 Juni 2021**

Penulis



**INTAN LARASATI**

**L200170091**

# **ANALISA PERBANDINGAN DATA MINING PADA KLASIFIKASI PENYAKIT JANTUNG MENGGUNAKAN ALGORITMA EXTREME LEARNING MACHINE (ELM) DAN K-NEAREST NEIGHBOR (K-NN)**

## **Abstrak**

Penyakit jantung merupakan salah satu jenis Penyakit Tidak Menular (PTM) yang mengakibatkan tingkat kematian yang cukup tinggi. Penyakit jantung disebabkan oleh beberapa faktor resiko diantaranya kebiasaan merokok, gaya hidup yang tidak sehat, tingginya kolesterol, hipertensi, dan diabetes. Berdasarkan fakta tersebut maka diperlukan algoritma yang tepat untuk mengklasifikasikan penyakit jantung sebagai salah satu upaya mencegah peningkatan angka kematian akibat penyakit jantung. Algoritma yang digunakan diharapkan dapat bekerja secara akurat dalam metode pengklasifikasian. Diantaranya terdapat dua algoritma yang digunakan yaitu algoritma Extreme Learning Machine (ELM) dan algoritma K-Nearest Neighbor (K-NN). Tujuannya adalah untuk membandingkan kedua algoritma tersebut, guna menentukan algoritma mana yang memiliki persentase ketepatan yang lebih tinggi dalam mengklasifikasi data penyakit jantung. Untuk mencapai tujuan dari penelitian maka dilakukan beberapa metode penelitian yaitu data preprocessing dengan tahapan data collection, data splitting dan normalisasi data dilanjutkan dengan metode algoritma ELM dan K-NN pada tahap data processing. Dari tahapan-tahapan yang telah dilakukan diperoleh hasil akhir algoritma Extreme Learning Machine (ELM) memperoleh nilai akurasi yang lebih besar yaitu sebesar 93,33% sedangkan algoritma K-NearestNeighbor (K-NN) memperoleh nilai akurasi sebesar 83,52%. Hal tersebut menunjukkan bahwa dalam penelitian ini algoritma Extreme Learning Machine (ELM) bekerja lebih maksimal dibandingkan algoritma K-Nearest Neighbor (K-NN) dalam klasifikasi data penyakit jantung.

**Kata Kunci:** Data Mining, Extreme Learning Machine, K-Nearest Neighbor, Penyakit Jantung

## **Abstract**

Heart disease is type of non communicable disease which results in a high mortality rate. Heart disease is caused by several risk factors including smoking, an unhealth lifestyle, high cholesterol, hypertension, and diabetes. Based on these facts, an appropriate algorithm is needed to classify heart disease as an effort to prevent an increase in the death rate from heart disease. The algorithm used is expected to work accurately in the classification method. Among them, there are two algorithm used, namely the Extreme Learning Machine (ELM) algorithm and the K-Nearest Neighbour (K-NN) algorithm. The aim is to compare the two algorithms, in order to determine which algorithm has the higher percentage of accuracy in classifying heart disease data. To achieve the objectives of the study, several research methods were carried out, namely data preprocessing with the data collection stage, data splitting and data normalization followed by the ELM and K-NN algorithm methods at the data processing stage. From the steps that have been carried out, the final result of the Extreme Learning Machine (ELM) algorithm obtained a greater accuracy value of 93.33%, while the K-Nearest Neighbour (K-NN) algorithm obtained an accuracy value of 83.52%. This shows that in this study

the Extreme Learning Machine (ELM) algorithm works more optimally than the K-Nearest Neighbour (K-NN) algorithm in the classification of heart disease data.

**Keywords:** Data Mining, Extreme Learning Machine, Heart Disease, K-Nearest Neighbour

## 1. PENDAHULUAN

Jumlah kasus kematian yang disebabkan oleh penyakit jantung kian hari kian meningkat. Mengutip data dari *World Health Organization* (WHO) saat ini telah lebih dari 17 juta jiwa kehilangan nyawa akibat penyakit jantung (Effendi and Widiastuti, 2014). Angka tersebut diprediksi akan terus mengalami peningkatan hingga mencapai 23,3 juta jiwa pada tahun 2030 (Bianto, Kusri and Sudarmawan, 2020). Sedangkan di Indonesia sendiri jumlah penderita penyakit jantung saat ini telah mencapai angka 2.784.064 orang, yang artinya 15 dari 1000 orang menderita penyakit tersebut (Effendi and Widiastuti, 2014). Penyakit jantung termasuk kedalam jenis Penyakit Tidak Menular (PTM). Penyakit jantung disebabkan oleh beberapa faktor risiko diantaranya kebiasaan merokok, gaya hidup yang tidak sehat, tingginya kolesterol, hipertensi, diabetes, dan stres (Syafitri Hidayatul AA, Yuita Arum S, 2018).

Sebagai upaya mencegah peningkatan angka kematian akibat penyakit jantung dapat dilakukan prediksi penyakit jantung yang terdapat pada manusia. Banyak teknologi yang bisa digunakan dalam teknik prediksi untuk mengelola data yang bisa membantu menentukan seseorang memiliki risiko penyakit jantung atau tidak. *Data mining* adalah salah satu teknologi yang paling banyak digunakan.

*Data mining* dikenal memiliki dua jenis model prediksi yaitu prediksi klasifikasi dan regresi. Klasifikasi *data mining* adalah suatu metode untuk memprediksi nilai dari sekelompok atribut dalam menggambarkan dan membedakan kelas data atau konsep yang bertujuan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui (Ardiyansyah, Rahayuningsih and Maulana, 2018). Dalam klasifikasi, algoritma yang sering digunakan adalah *naïve&bayes*, *neural network*, C4.5, SVM, K-NN dan lain sebagainya.

Pada penelitian sebelumnya mengenai klasifikasi penyakit jantung yang menggunakan gabungan lima algoritma yaitu *decision tree*, *k-nearest neighbor*, *Naïve bayes*, *random forest*, dan *decision stump* diperoleh nilai akurasi tertinggi sebesar

80.38% dan menunjukkan bahwa algoritma *random forest* dan *decision stump* memiliki performa terbaik sedangkan algoritma K-NN memiliki performa yang kurang baik (Annisa, 2019). Disisi lain (Khadijah and Kusumaningrum, 2019) dalam penelitiannya yang menggunakan algoritma ELM, SVM, dan K-NN untuk klasifikasi kanker payudara mengungkapkan bahwa ketiga algoritma tersebut memberikan hasil yang hampir sama atau memiliki nilai yang berdekatan. Terdapat pula penelitian lain yang menerapkan algoritma *Extreme Learning Machine* (ELM) pada klasifikasi risiko gagal ginjal kronis dengan hasil nilai akurasi yang terbaik adalah sebesar 99,13% (Prenky *et al.*, 2018).

Berdasarkan penelitian yang telah dilakukan sebelumnya serta masih banyak lagi penelitian sejenis lainnya. Maka pada penelitian ini akan dilakukan analisa komparasi antara algoritma *Extreme Learning Machine* (ELM) dan *K-Nearest Neighbor* (K-NN) pada klasifikasi penyakit jantung. *Extreme Learning Machine* (ELM) merupakan algoritma yang diusulkan sebagai pembelajaran baru untuk jaringan *feedforward* dan sering kali disebut sebagai *Single Layer Feedforward Neural Networks* (SLFNs) (Rubiolo, Milone and Stegmayer, 2018). Model jaringan SLFN adalah jaringan tiga lapis atau *three-layer network* dimana  $n$  sebagai *node input*,  $L$  sebagai *node* komputasi dilapisan tersembunyi dan  $m$  sebagai *node output* (Shen *et al.*, 2018). K-NN adalah algoritma yang tujuannya untuk mengelompokkan data baru berdasarkan atribut dan data *training* (Prajarini *et al.*, 2016). Penelitian ini bertujuan untuk menentukan algoritma yang memiliki nilai akurasi yang lebih baik dalam klasifikasi penyakit jantung.

## 2. METODE

Pemrosesan data pada penelitian ini akan dilakukan dengan menggunakan bahasa pemrograman Python dalam *software* Jupyter Notebook. Pada penelitian ini akan digunakan beberapa metode sebagai berikut :

## 2.1 Data Collection

Penelitian ini menggunakan dataset penyakit jantung yang tersedia secara publik pada *Kaggle Datasets*. Dataset yang digunakan terakhir kali diperbarui pada tahun 2018 dan dibuat oleh Andras Janosi, M.D dari Hungarian Institute of Budapest Cardiology, Willem Steinbrunn, M.D dari University Hospital, Zurich, Switzerland, Matthias Pfisterer, M.D. dari University Hospital, Basel, Switzerland, dan Robert Detrano, M.D., Ph.D. dari V.A. Medical Center, Long Beach and Cleveland Clinic Foundation. Dataset mempunyai 14 atribut terdiri dari 303 data. Atribut-atribut kemudian diberi variabel X dan Y, dimana X merupakan variabel *input*, dan Y merupakan variabel target variabel dan atribut dataset akan ditunjukkan dalam Tabel 1.

Tabel 1. Variabel dan Atribut Penderita Penyakit Jantung

Variabel	Atribut	Deskripsi
X1	<i>Age</i>	Usia dalam tahun
X2	<i>Sex</i>	1= laki-laki ; 0 = wanita
X3	<i>Cp</i>	<i>Chest Pain Type</i> : Value 1: typical angina; Value 2: atypical angina; Value 3: non-anginal pain; Value 4: asymptomatic
X4	<i>Trestbps</i>	Tekanan darah dalam mm Hg saat masuk ke rumah sakit
X5	<i>Chol</i>	Serum kolestoral dalam mg / dl
X6	<i>Fbs</i>	Gula darah > 120mg / dl 1 = benar ; 0 = salah
X7	<i>Restecg</i>	Hasil elektrokardiografi istirahat - Nilai 0 : normal - Nilai 1 : memiliki kelainan gelombang ST-T (inversi gelombang T dan / atau elevasi atau depresi ST > 0,05 mV) - Nilai 2: menunjukkan kemungkinan atau pasti hipertrofi
X8	<i>Thalach</i>	Detak jantung maksimum tercapai
X9	<i>Exang</i>	Angina yang diinduksi latihan (1 = ya ; 0 = tidak)
X10	<i>Oldpeak</i>	Depresi ST yang disebabkan oleh olahraga relatif terhadap istirahat
X11	<i>Slope</i>	Kemiringan segmen ST latihan puncak Nilai 1 : miring ke atas ; Nilai 2 : datar ; Nilai 3 : landai
X12	<i>Ca</i>	Jumlah kapal besar (0-3) yang diwarnai dengan tepung
X13	<i>Thal</i>	3 = normal ; 6 = cacat tetap ; 7 = cacat yang dapat diperbaiki
Y	<i>Num</i>	Diagnosis penyakit jantung (status penyakit)



		angiografi) Nilai 0: penyempitan diameter < 50% ; Nilai 1: penyempitan diameter > 50%
--	--	---

## 2.2 Data Splitting

*Data splitting* merupakan metode yang digunakan untuk memvalidasi silang dengan cara membagi *dataset* menjadi dua yaitu satu untuk mengkalibrasi dan yang lainnya untuk validasi (De Souza, 2019). Metode yang akan digunakan pada tahapan *data splitting* adalah metode *cross validation*, yaitu teknik untuk membuktikan ketepatan suatu model yang dibangun didasarkan pada *dataset* tertentu (Supartini, Sukarsa and Srinadi, 2017). Pada penelitian ini, dilakukan *splitting* yang dibagi menjadi 70% data latih dan 30% data uji hal ini didasarkan pada penelitian sebelumnya yaitu klasifikasi terhadap penyakit kanker payudara menggunakan algoritma *Naïve Bayes* yang melakukan pengujian terhadap persentase *splitting* yang berbeda-beda kemudian dari penelitian tersebut diketahui bahwa dengan persentase *splitting* 70% dan 30% diperoleh hasil akurasi yang tinggi (Oktavianto and Handri, 2020).

## 2.3 Data Normalisasi

Untuk normalisasi data, metode yang digunakan adalah metode *Min-Max Normalization*. *Min-Max normalization* merupakan metode normalisasi dengan melakukan transformasi linier terhadap data asli sehingga menghasilkan keseimbangan nilai perbandingan antar data saat sebelum dan sesudah proses (Nasution, Khotimah and Chamidah, 2019). Metode ini dapat menggunakan rumus Persamaan 1.

$$x_{normal} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

## 2.4 Extreme Learning Machine (ELM)

ELM merupakan algoritma jaringan saraf tiruan *feed-forward* yang diusulkan oleh Huang. ELM memiliki fitur-fitur yang cukup berbeda dengan jaringan saraf *feed-forward* lain yang biasanya menggunakan algoritma pembelajaran berbasis gradien (Pangaribuan, 2016). Berikut adalah fitur-fitur ELM :

- a. Dibandingkan dengan pelatihan jaringan saraf *feed-forward* yang menggunakan algoritma pembelajaran klasik yang memakan waktu cukup lama, fase pembelajaran ELM terhitung sangat cepat bahkan dapat diselesaikan dalam hitungan detik.

- b. Kinerja generalisasi ELM lebih baik dibandingkan pembelajaran dengan basis gradien.
- c. ELM jauh lebih sederhana dari algoritma pembelajaran jaringan saraf *feed-forward* lainnya sehingga bisa meminimalisir permasalahan yang biasa terjadi pada algoritma pembelajaran klasik.
- d. ELM dapat digunakan untuk melatih fungsi SLFNs dengan berbagai macam fungsi aktivasi yang tidak terdiferensiasi, tidak seperti algoritma berbasis gradien yang hanya bekerja untuk fungsi aktivasi terdiferensiasi.

Langkah-langkah perhitungan metode ELM dibagi menjadi 2 proses, yaitu proses *training* dan proses *testing*.

#### 2.4.1 Proses *Training*

Proses perhitungan ELM dalam melakukan proses *training* adalah sebagai berikut :

- a. Inisialisasi *Input Weight* ( $W_{jk}$ ) dan Bias ( $b$ ) Secara *Random*

Langkah pertama yang dilakukan pada proses *training* adalah melakukan inisialisasi *input weight* dengan ukuran matriks  $j \times k$  dan bias dengan ukuran  $1 \times j$  dimana  $j$  merupakan banyak *hidden node* dan  $k$  adalah banyak *input node*.

- b. Menghitung Matriks *Output Hidden Layer* ( $H$ )

*Output hidden layer* telah diaktivasi menggunakan fungsi aktivasi sigmoid biner yang digunakan untuk memetakan nilai matriks *output hidden layer* pada interval 0 sampai 1.

Perhitungan *output hidden layer* ditunjukkan pada Persamaan 2.

$$H = 1 / (1 + \exp(-(X \cdot W^T + \text{ones}(N_{\text{train}}, 1) \times b))) \quad (2)$$

Keterangan :

$H$  = *Output Hidden Layer*

$\exp$  = Eksponensial

$X$  = Matriks Data *Input*

$W^T$  = Matriks *Transpose Input Weight*

$\text{ones}(N_{\text{train}}, 1)$  = Matriks yang bernilai 1 dengan ukuran baris sama dengan jumlah data *training* dan 1 kolom.

$b$  = Bias

- c. Menghitung Matriks *Moore-Penrose Pseudo Inverse*

Matriks *Moore-Penrose Psudo Invers* adalah matriks yang diperoleh dari perkalian matriks *inverse* dan *transpose output hidden layer*. Perhitungan Matriks *Moore-Penrose Pseudo Invers* ditunjukkan pada Persamaan 3.

$$H^+ = (H^+ \cdot H)^{-1} \cdot H^T \quad (3)$$

Keterangan :

$H^+$  = Matriks *Moore-Panrose Pseudo Invers*

$H$  = Matriks *Output Hidden Layer*

$H^T$  = *Transpose Matriks Output Hidden Layer*

d. Menghitung *Output Weight / Output ( $\beta$ )*

*Output weight / Output ( $\beta$ )* merupakan bobot yang dihasilkan oleh *hidden layer* dengan *output layer*. Perhitungan *output weight* terdapat pada Persamaan 4.

$$\beta = H^+ \cdot Y \quad (4)$$

Keterangan :

$\beta$  = *Output weight Hidden Layer*

$H^+$  = Matriks *Moore-Panrose Pseudo Invers*

$Y$  = Matriks *Data Output* atau *Target*

e. Menghitung Hasil Prediksi ( $\hat{Y}$ )

Hasil prediksi ELM merupakan proses perkalian antara matriks *output hidden layer* dengan *output weight hidden layer*. Perhitungan hasil prediksi ELM ditunjukkan pada Persamaan 5.

$$\hat{Y} = H \cdot \beta \quad (5)$$

Keterangan :

$\hat{Y}$  = Hasil Prediksi

$H$  = Matriks *Output Hidden Layer*

$\beta$  = Matriks *Output Weight Hidden Layer*

#### 2.4.2 Proses *Testing*

Berbeda dengan proses *training* yang digunakan untuk membuat pola model ELM, proses *testing* digunakan untuk mengevaluasi kemampuan ELM sebagai *forecasting tool*. Langkah-langkah proses *testing* sama dengan yang dilakukan pada proses *training*, tetapi pada proses *testing* tidak perlu menghitung *input weight*, bias, dan *output weight*. *Input weight*, bias, dan *output weight* pada proses *testing* menggunakan bobot yang telah dihitung dari proses *training*.

## 2.5 K-NN

Algoritma K-NN adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan  $k$  buah data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Nilai  $k$  tersebut.

K-NN bekerja secara sederhana, pada fase *training* algoritma akan menyimpan seluruh objek dari data *training* dan mengklasifikasikan objek-objek baru berdasarkan kemiripan (Yustanti, 2012). Kemudian pada fase klasifikasi, objek dengan fitur yang sama akan dihitung untuk digunakan sebagai data *testing* yang klasifikasinya tidak diketahui (Yustanti, 2012). Objek baru diklasifikasikan berdasarkan jarak objek baru ke beberapa objek terdekat. Berikut adalah langkah-langkah untuk melakukan perhitungan menggunakan algoritma K-NN:

- a. *Input* data *training*, data *testing*, serta nilai  $k$ .
- b. Menghitung jarak antara data *training* dan data *testing*.
- c. Dilakukan penyortiran dengan dasar kedekatan jarak serta mendapatkan data *training* teratas sebanyak nilai  $k$ .
- d. Menentukan kelas klasifikasi berdasarkan kelas yang setara dan memiliki nilai terbanyak.

Untuk menghitung jarak digunakan fungsi *Euclidean Distance* pada Persamaan 6.

$$Euclidean = \sqrt{\sum_{i=1}^k (y_{2i} - y_{1i})^2} \quad (6)$$

Keterangan :

<i>Euclidean</i>	= jarak
$y_2$	= data latih
$y_1$	= data uji
$i$	= variable data
$k$	= dimensi data

## 2.6 Evaluation Model

Evaluasi dilakukan guna mengetahui akurasi dari algoritma yang digunakan (Utami, 2017). Akurasi adalah skala dari jumlah prediksi *true* dari keseluruhan data. Untuk menghitung nilai akurasi digunakan rumus persamaan seperti pada Persamaan 7.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

### 3. HASIL DAN PEMBAHASAN

Penelitian ini diawali dengan proses pengumpulan *dataset* penyakit jantung, *dataset* yang digunakan diperoleh dari *Kaggle Datasets* sejumlah 303 data dengan 14 variabel yang terdiri dari 13 variabel input dan 1 variabel target. Berdasarkan data yang telah diperoleh, kemudian dilakukan perbandingan klasifikasi dengan menggunakan algoritma *Extreme Learning Machine* (ELM) dan algoritma *K-Nearest Neighbor* (K-NN).

Dataset tersebut terbagi secara acak menjadi 70% *data training* dan 30% *data testing*. Selanjutnya dilakukan normalisasi terhadap *dataset* penyakit jantung karena banyak data yang memiliki rentang berbeda. Dataset ditransformasi dengan metode *Min-Max Normalization* dengan mengolah nilai minimum dan maksimum dari setiap atribut.

Tabel 2. Data Sebelum di Normalisasi

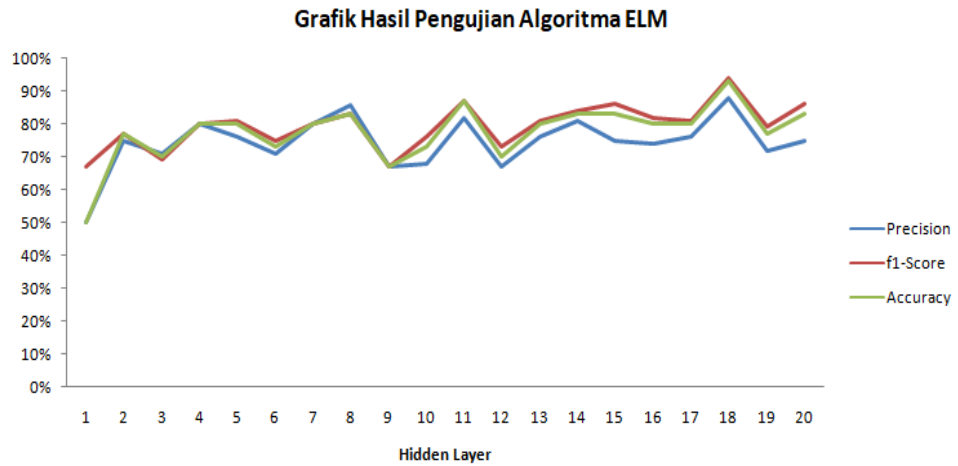
	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2

Dari Tabel 2 dapat diketahui bahwa atribut-atribut pada data penyakit jantung yang digunakan dalam penelitian ini memiliki rentang nilai yang cukup besar. Karenanya harus dilakukan normalisasi untuk memperoleh rentang dengan nilai yang lebih kecil dengan bobot yang sama, sehingga dapat memudahkan, meningkatkan kualitas data, dan meminimalisir kesalahan dalam proses *learning* terhadap model pembelajaran (Yustanti, 2012). Dalam hal ini normalisasi dilakukan dengan metode *Min-Max Normalization* dengan rentang nilai 0 – 1 sehingga dihasilkan data seperti pada Tabel 3.

Tabel 3. Data Setelah di Normalisasi

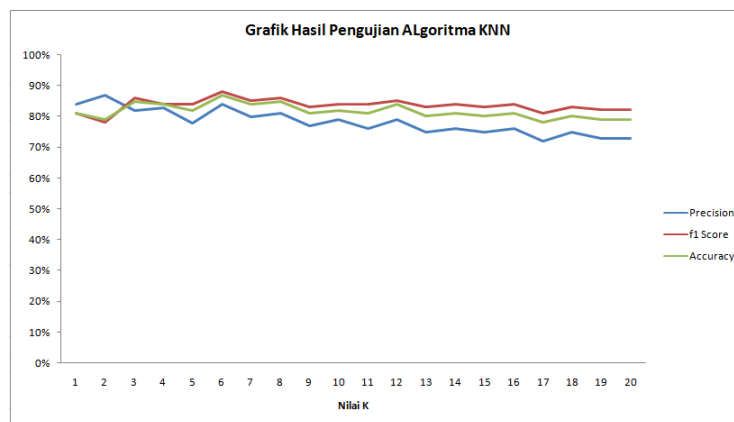
	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal
0	0.708333	1.0	1.000000	0.481132	0.244292	1.0	0.0	0.603053	0.0	0.370968	0.0	0.0	0.333333
1	0.166667	1.0	0.666667	0.339623	0.283105	0.0	0.5	0.885496	0.0	0.564516	0.0	0.0	0.666667
2	0.250000	0.0	0.333333	0.339623	0.178082	0.0	0.0	0.770992	0.0	0.225806	1.0	0.0	0.666667
3	0.562500	1.0	0.333333	0.245283	0.251142	0.0	0.5	0.816794	0.0	0.129032	1.0	0.0	0.666667
4	0.583333	0.0	0.000000	0.245283	0.520548	0.0	0.5	0.702290	1.0	0.096774	1.0	0.0	0.666667

Tahap selanjutnya dilakukan pemrosesan *dataset* yang diolah dengan bahasa pemrograman *Python* dengan *tools* Jupyter Notebook dalam aplikasi Anaconda Navigator. Pemrosesan data tersebut kemudian memperoleh hasil rata-rata nilai akurasi dari algoritma *Extreme Learning Machine* (ELM) dan *K-Nearest Neighbor* (K-NN). Kinerja dari algoritma ELM dipengaruhi oleh Hidden Neuron, grafik pada Gambar 1 adalah grafik hasil pengujian pengaruh jumlah hidden neuron.



Gambar 1. Grafik Hasil Pengujian Algoritma ELM

Dari pengujian algoritma ELM diperoleh hasil rata-rata akurasi sebesar 93,33%, rata-rata *precision* 88% dan rata-rata f1-score 94%. Dari grafik pada Gambar 3 dapat diketahui bahwa pada penelitian ini algoritma bekerja lebih optimal dengan 18 *hidden layer*. Sedangkan algoritma *K-Nearest Neighbor* (K-NN) mendapatkan rata-rata nilai akurasi sebesar 83.52%, rata-rata *precision* 78,25%, dan rata-rata f1-score 83,5%. Kinerja dari algoritma K-NN sangatlah dipengaruhi oleh nilai *k*. Grafik pada Gambar 2 menunjukkan hasil dari pengujian pengaruh nilai *k*.



Gambar 2. Grafik Pengujian Algoritma K-NN

Pengujian nilai  $k$  dilakukan dengan rentang nilai 1-20. Hasil yang diperoleh dari pengujian terhadap nilai  $k$  pada penelitian ini adalah algoritma K-NN bekerja maksimal dengan nilai  $k = 3$ . Grafik pada Gambar 2 juga menunjukkan bahwa *score accuracy* mengalami penurunan signifikan pada nilai  $k$  sebanyak 6, semakin banyak jumlah nilai  $k$  yang digunakan justru membuat kinerja dari algoritma K-NN tidak berjalan maksimal.

#### 4. PENUTUP

Penelitian ini dilakukan untuk membandingkan nilai akurasi dari kedua algoritma yaitu *Extreme Learning Machine* (ELM) dan *K-Nearest Neighbor* (K-NN). Hasil penelitian dari pengujian kedua algoritma untuk klasifikasi penyakit jantung menghasilkan performa terbaik dengan rata-rata nilai akurasi sebesar 93,33% adalah algoritma *Extreme Learning Machine* (ELM). Sedangkan algoritma *K-Nearest Neighbor* (K-NN) memperoleh rata-rata nilai akurasi lebih rendah sebesar 83,52%. Berdasarkan hasil penelitian yang telah dilaksanakan, masih terdapat kekurangan. Terdapat beberapa rekomendasi yang dapat dilakukan untuk penelitian selanjutnya yaitu mengembangkan algoritma pada jenis dataset yang lain serta dapat menggunakan algoritma klasifikasi sejenis yang lain. kemudian dari hasil yang diperoleh kemudian dapat digunakan untuk kebutuhan pengembangan sistem informasi dalam klasifikasi dan prediksi penyakit jantung.

#### DAFTAR PUSTAKA

- Annisa, R. (2019) ‘Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung’, *Jurnal Teknik Informatika Kaputama (JTIK)*, 3(1), pp. 22–28. Available at: <http://jurnal.kaputama.ac.id/index.php/JTIK/article/view/141>.
- Ardiyansyah, Rahayuningsih, P. A. and Maulana, R. (2018) ‘Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk Dataset Blogger Dengan Rapid Miner’, *Jurnal Khatulistiwa Informatika*, 6(1), pp. 20–28. doi: 10.31294/jki.v6i1.3799.g2437.
- Bianto, M. A., Kusriani, K. and Sudarmawan, S. (2020) ‘Perancangan Sistem Klasifikasi Penyakit Jantung Menggunakan Naïve Bayes’, *Creative Information Technology Journal*, 6(1), p. 75. doi: 10.24076/citec.2019v6i1.231.
- Effendi, N. and Widiastuti, H. (2014) ‘Kebiasaan Olah Raga Terhadap Kejadian

- Penyakit Jantung', *Jurnal Kesehatan*, 7(2), pp. 353–360. doi: 10.24252/kesehatan.v7i2.54.
- Khadijah, K. and Kusumaningrum, R. (2019) 'Ensemble Classifier untuk Klasifikasi Kanker Payudara', *It Journal Research and Development*, 4(1), pp. 61–71. doi: 10.25299/itjrd.2019.vol4(1).3540.
- Nasution, D. A., Khotimah, H. H. and Chamidah, N. (2019) 'Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN', *Computer Engineering, Science and System Journal*, 4(1), p. 78. doi: 10.24114/cess.v4i1.11458.
- Oktavianto, H. and Handri, R. P. (2020) 'Analisis Klasifikasi Kanker Payudara Menggunakan Algoritma Naive Bayes', *INFORMAL: Informatics Journal*, 4(3), p. 117. doi: 10.19184/isj.v4i3.14170.
- Pangaribuan, J. J. (2016) 'Mendiagnosis Penyakit Diabetes Melitus Dengan Menggunakan Metode Extreme Learning Machine', *Jurnal ISD*, 2(2), pp. 2528–5114.
- Pertiwi, M. W., Adiwisatra, M. F. and Supriadi, D. (2019) 'Analisa Komparasi Menggunakan 5 Metode Data Mining dalam Klasifikasi Persentase Wanita Sudah menikah di Usia 15-49 yang Memakai Alat KB (Keluarga Berencana)', *Jurnal Khatulistiwa Informatika*, 7(1), pp. 37–42. doi: 10.31294/jki.v7i1.5741.
- Prajarini, D. *et al.* (2016) 'Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Kulit', *Informatics Journal*, 1(3), p. 137.
- Prenky, D. *et al.* (2018) 'Klasifikasi Risiko Gagal Ginjal Kronis Menggunakan Extreme Learning Machine', *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 2(11), pp. 5220–5228.
- Rivki, M. and Bachtiar, A. M. (2017) 'Implementasi Algoritma K-Nearest Neighbor Dalam Pengklasifikasian Follower Twitter Yang Menggunakan Bahasa Indonesia', *Jurnal Sistem Informasi*, 13(1), p. 31. doi: 10.21609/jsi.v13i1.500.
- Rubiolo, M., Milone, D. H. and Stegmayer, G. (2018) 'Extreme Learning Machines for Reverse Engineering of Gene Regulatory Networks from Expression Time Series', *Bioinformatics*, 34(7), pp. 1253–1260. doi: 10.1093/bioinformatics/btx730.
- Shen, Y. *et al.* (2018) 'An Ensemble Method Based on Selection Using Bat Algorithm for Intrusion Detection', *Computer Journal*, 61(4), pp. 526–538. doi: 10.1093/comjnl/bxx101.
- De Souza, K. (2019) 'Two cross-validation techniques to comprehensively characterize global horizontal irradiation regression models: Single data-splitting is insufficient', *Journal of Renewable and Sustainable Energy*, 11(6), p. 063702. doi: 10.1063/1.5116642.
- Supartini, I. A. M., Sukarsa, I. K. G. and Srinadi, I. G. A. M. (2017) 'Analisis



Diskriminan Pada Klasifikasi Desa Di Kabupaten Tabanan Menggunakan Metode K-Fold Cross Validation', *E-Jurnal Matematika*, 6(2), p. 106. doi: 10.24843/mtk.2017.v06.i02.p154.

- Syafitri Hidayatul AA, Yuita Arum S, A. A. (2018) 'Seleksi Fitur Information Gain untuk Klasifikasi Penyakit Jantung Menggunakan Kombinasi Metode K-Nearest Neighbor dan Naïve Bayes', *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(9), pp. 2546–2554.
- Utami, L. A. (2017) 'Analisis Sentimen Opini Publik Berita Kebakaran Hutan Melalui Komparasi Algoritma Support Vector Machine Dan K-Nearest Neighbor Berbasis Particle Swarm Optimization', 13(1), pp. 103–112.
- Yustanti, W. (2012) 'Algoritma K-Nearest Neighbour untuk Memprediksi Harga Jual Tanah', 9(1), pp. 57–68.